COMMENTS OF THE
UNITED STATES INTERNET SERVICE PROVIDER ASSOCIATION

APPENDIX A

TECHNICAL ISSUES REGARDING CALEA COMPLIANCE FOR VOICE OVER INTERNET PROTOCOL AND OTHER IP-BASED SERVICES

I. INTRODUCTION

Applying CALEA to services based on Internet Protocol ("IP") presents numerous and significant technical challenges, such as handling requests for call-identifying information ("CII") that is not "reasonably available" to service providers as defined by section 103 of CALEA. The complex technical issues described in this appendix also make it critical for industry experts to lead the development of standards for delivery of call content and CII pursuant to CALEA, rather than for such standards to be mandated by law enforcement or the Federal Communications Commission.

This appendix describes various technical difficulties with extracting CII and call content on IP-based networks, particularly where multiple service providers are involved with a particular service (*e.g.*, where voice over IP ("VoIP") service is offered by one service provider over the broadband access service of another). Many of these difficulties are associated with the technical and practical hurdles faced by service providers if they have to "break open" packets – examine parts of the IP packet that they would not normally examine as part of their regular operations – to extract CII or call content.

This technical appendix focuses on challenges associated with provision of call content and CII for VoIP service, but also addresses various issues for IP-based services more generally. The appendix begins with an explanation of VoIP service, including the relevant protocols and the differences from circuit-switched communications. Next, the appendix separately addresses technical challenges for broadband access providers and VoIP providers.

II. VOICE OVER IP ("VOIP")

VoIP utilizes a complex and developing suite of protocols to transmit voice content over the Internet or other IP networks. In general, IP networks are designed to transmit packets in a consistent manner regardless of their content, with much of the "intelligence" that processes the content located in hardware or software at the network edge. This content-independent network architecture is different than the architecture of circuit-mode telephone networks, which are designed and optimized for the specific purpose of transmitting voice calls.[1] Under the traditional IP architecture, it is quite possible that a VoIP service provider supporting a VoIP call may have only a limited role in setting up a call, or that a broadband access provider transmitting the call content may be unaware that a VoIP call is even taking place on its network. Before exploring this point, though, it is first necessary to discuss how VoIP calls are executed.

VoIP relies on one set of protocols for call signaling (*e.g.*, initiating/terminating a call, negotiating the voice encoding mechanisms to be used) and another set for the transmission of call content. This allows call signaling information and call content to be handled independently, often over different network paths. While it is possible for signaling information and digitized call content to follow identical paths from sender to receiver, it is likely that the signaling data would pass through one or more proxy or redirect servers while the call content takes a more direct path. This separation of signaling information from call content is crucial for understanding where CII may be located.

---

[1] Some "enhanced" IP architectures are optimized for VoIP and other real-time applications. Also described as "business-class," these networks seek to provide higher levels of reliability, security and quality of service.

A. VoIP Content Protocols

In almost all VoIP implementations, the Real-time Transport Protocol ("RTP") is used to transmit the digitized voice content associated with a VoIP call. RTP defines a standardized format for transporting audio and video content over IP networks that can be used in numerous applications, including streaming video and audio, and videoconferencing.[2] In VoIP applications, RTP packets contain the digitized voice content of a call, which may be encoded using one of many available coding schemes.

RTP is built on top of the User Datagram Protocol ("UDP"), which is itself built on top of the IP protocol. Thus, an RTP packet includes three separate layers of headers (IP, UDP and RTP), each with information relevant to the manner in which the packet should be processed. The IP header contains the source and destination IP addresses which are used by routers to forward the packet to its destination; the UDP header contains the source and destination ports which are used by the receiving computer to choose the appropriate application to process the packet; and the RTP header includes information necessary for the destination application to reconstruct the original voice sample (or other audio/video content). For example, the RTP header contains a field indicating the payload type for a given packet which identifies the coding scheme used to produce the data enclosed in the packet payload. The RTP header also contains a sequence number that allows the receiver to detect packet loss and/or packets arriving in the wrong order, and a timestamp that allows for synchronized playback.[3] In general, in order to route packets to their destination, broadband access (or Internet backbone) service providers

[2] H. Schulzrinne, et al., RFC 3550 - RTP: A Transport Protocol for Real-Time Applications (2003), *available at* www.ietf.org/rfc/rfc3550.txt ("RFC 3550 - RTP").

[3] DANIEL COLLINS, *CARRIER GRADE VOICE OVER IP* 53 (2003).

examine only the information contained in the IP header of a given packet.[4] Therefore, such a service provider would ordinarily have no reason to know that an IP packet contains RTP content. Broadband access providers would have to modify their networks to examine information beyond a packet's IP header.

Adding to the complexity, many VoIP applications that use RTP also utilize a companion protocol – the RTP Control Protocol ("RTCP") – to convey control information between session participants and provide feedback on the quality of data distribution. RTCP monitors performance factors related to the delivery of RTP packets, including variation in delivery delay (known as jitter), cumulative number of packets dropped in a session, round-trip time for a packet to travel to a destination and back, etc.[5] Using these data, RTCP is able to provide critical quality of service ("QoS") feedback that can be used by session participants or separate session monitors to detect and potentially correct distribution problems.[6]

B. VoIP Signaling Protocols

VoIP utilizes a separate protocol for the signaling portion of a call. Session Initiation Protocol ("SIP") is considered the leading signaling protocol for VoIP.[7] Created by the Internet Engineering Task Force ("IETF") as a simpler alternative to the ITU-T's H.323 signaling protocol, SIP is a standard for initiating, managing, and terminating communication sessions

---

[4] Some broadband access providers utilize bandwidth management tools to improve the performance of specific applications (*e.g.*, VoIP), control user costs and optimize network efficiency. Such tools may examine packets beyond just the IP header.

[5] *See* RFC 3550 - RTP.

[6] *Id.*

[7] J. Rosenberg, et al., RFC 3261 - SIP: Session Initiation Protocol (2002), *available at* www.ietf.org/rfc/rfc3261.txt ("RFC 3261 - SIP").

between one or more VoIP network endpoints (which may be implemented in hardware or software). In developing SIP, IETF sought to replicate many of the call processing functions and features present in the public switched telephone network ("PSTN"), such as dialing a number, causing a phone to ring and hearing a busy signal; and SIP incorporates many of the call processing features of Signaling System 7 ("SS7") used on circuit-switched networks. However, because of the significant differences between circuit-switched and IP-based networks, SIP is very different from SS7, and implements call processing in an entirely different manner than on the PSTN. Whereas PSTN/SS7 networks are characterized by a highly complex central network architecture specifically designed to support voice communications, SIP is a "peer-to-peer" protocol that allows numerous kinds of media to be communicated over a relatively simple core network.[8] The protocol is described as peer-to-peer because two SIP endpoints can communicate without any intervening SIP infrastructure.[9]

SIP may be used on top of any datagram or stream transport layer protocol – the most common of which are the Transmission Control Protocol ("TCP") and UDP – which in turn is implemented on top of the IP protocol.[10] Thus, like RTP packets, a SIP packet includes three separate headers (IP; UDP, TCP, etc.; and SIP), each with information relevant to the way the packet should be processed. The IP header of a SIP packet contains the source and destination IP addresses; the header of the transport layer protocol contains the source and destination ports;

[8] RFC 3261 - SIP.

[9] As a practical matter, SIP requires intermediate infrastructure in the form of proxy and registrar servers; and these are typically offered by VoIP service providers.

[10] UDP is generally regarded as the preferred transport layer protocol, because of its superior performance and scalability. When a message is too large for transport via UDP, TCP may be used. In some instances, SIP may be used on top of the Stream Control Transmission Protocol ("SCTP") instead of TCP.

and the SIP header describes the packet as either a request from a client to a server or a response from a server to a client (also called a status message). The SIP header also contains other fields to provide additional information about the message or to identify how the message should be handled.[11] These header fields may include information such as To, From, Subject, Content-Encoding, Date or Priority. The headers included in a given packet are dependent upon the message type of the packet. For example, a Subject header field containing user-defined text may be included in an INVITE message sent from one user to another to initiate a call session, but would not be present in a BYE message terminating a session. As with RTP, where a broadband access (or Internet backbone) service provider transmits SIP signaling information, its routers and other facilities only process the IP header in order to route the SIP packets to their destination; and the service provider would ordinarily have no reason to know that such IP packets contain SIP content.

Although there is an increasing degree of standardization around SIP, VoIP communications can also be controlled using a variety of other protocols: H.323, the ITU-T signaling protocol for multimedia communications over packet-switched networks; Skinny Client Control Protocol ("SCCN"), a semi-proprietary VoIP terminal control protocol defined by Cisco; Media Gateway Control Protocol ("MGCP"), used with SIP in a softswitch architecture; and others. This multiplicity of protocols creates even greater difficulties for the broadband access providers attempting to retrieve VoIP call content and CII, since such a transport provider will usually not know which protocols the customer's VoIP provider is using.[12]

---

[11] DANIEL COLLINS, CARRIER GRADE VOICE OVER IP 176 (2003).

[12] Indeed, persons or entities seeking to secure their communications (*i.e.*, those in which law enforcement is most interested) may intentionally choose a non-standard "flavor" of VoIP.

III. TECHNICAL CHALLENGES FOR BROADBAND ACCESS SERVICE PROVIDERS

Broadband access service providers generally have access to the source and destination IP addresses of the packets that they transmit on behalf of their customers, and it is likely to be feasible for such service providers to deliver the IP packet stream to law enforcement. Indeed, various existing CALEA standards (*e.g.*, the J-STD-025-A standard for packet mode networks) provide for such an approach. However, broadband access service providers typically do not process packets at a layer higher than the IP layer, or operate hardware or software that provides the ability to "break open" packets and examine content at higher network layers.[13] Accordingly, broadband access service providers would have to modify their networks significantly in order to identify and extract VoIP traffic transmitted over their infrastructure by another service provider, or VoIP traffic generated using functionality based solely on software or hardware supplied by the end customer. The need to modify networks is key to the question of whether CII is "reasonably available," addressed in detail in the main body of these comments.

This section discusses technical aspects of the process of "breaking open" packets, and then presents certain technical solutions that may assist in doing so. However, given the significant technical complexities of available solutions, it may be simpler and more cost effective for law enforcement to simply request the SIP signaling information and RTP call content directly from the VoIP provider that processes such information, rather than from the broadband access provider.

[13] Packet protocols and information at lower network layers (*e.g.*, the physical layer or data link layer) are more likely to be available to a broadband access service provider, but are fairly unlikely to be of interest to law enforcement.

A. The Process of "Breaking Open" Packets

1. Call-Identifying Information

As described in section II above, much of the CII for VoIP calls – *e.g.*, signaling information related to setting up, maintaining and terminating such calls – is transmitted separately from call content, often using SIP. This signaling information is located in SIP packets encapsulated with a UDP or TCP header, and then with an IP header. The only way to differentiate SIP packets from non-SIP packets is to identify that the packet is sent to or from port 5060, which is generally used for SIP.[14] Thus, in order to route SIP packets to law enforcement, broadband providers that normally do no more than route traffic based on IP header information would need to at least (1) isolate all packets sent to or from the targeted IP address, (2) "break open" the isolated packets in order to view the source and destination ports located in the UDP or TCP header, and (3) filter only those sent to or from port 5060. This process would be even more complicated where a protocol other than SIP is used for signaling, since other protocols generally do not use port 5060. For VoIP signaling using other protocols, the broadband service provider would have to know which port numbers are associated with those protocols and filter accordingly (assuming that those protocols use well-defined ports), or would have to examine all packets to identify potential signaling information.

Even assuming SIP is used, the above process would not disclose CII for the VoIP calls in question. To obtain CII, the broadband access service provider would need to (1) "break open" SIP packets, (2) determine which SIP packets are associated with VoIP (and not some other kind of SIP-based service), and (3) examine and parse the content of the SIP messages.

[14] In certain circumstances (*e.g.*, custom-configured private networks), a port other than 5060 may be used for SIP packets. This significantly increases the difficulties involved in identifying and extracting SIP packets in such circumstances.

Assuming this could be done, the CII would then have to be correlated with the associated VoIP content, which would typically be transmitted separately in an RTP packet stream. This may be difficult to do without active cooperation from the VoIP provider.

It is important to note that some CII required under CALEA for circuit-switched telephony may not be transmitted via SIP packets. For example, post-cut-through digits for the dialed digit extraction ("DDE") capability would be encoded as tones in RTP packets. In this situation, broadband providers would need to (1) identify and extract the RTP packets sent to/from the targeted individual (as above, this would require "breaking open" two layers of packet headers – first the IP header, then the TCP or UDP header), (2) reconstruct the RTP packet stream associated with a particular call (which may require access to SIP data as well), (3) examine the RTP headers to determine which encoding scheme was used, (4) decode the digitized content in the RTP packets, and (5) determine whether or not any of the decoded content is dialed digits.

2. Call Content

For a broadband access provider, the call content that it transmits as a business is an IP packet stream, which can be used for providing any of a wide variety of Internet-based services. The provider typically has no way of knowing which of these packets contain VoIP traffic. If law enforcement were to obtain a Title III intercept order for the full packet stream, it is relatively straightforward for a broadband access provider to deliver the entire IP stream (including the VoIP call content). However, there would be serious technical difficulties if law enforcement wants the broadband provider to isolate the VoIP call content – which is either (a) call content of services provided by a different service provider or (b) generated by end-user software or hardware.

As explained in section II above, VoIP call content is transported using the RTP protocol, separate from the signaling information that is typically transported using SIP. This means that encoded voice data is included in the payload of packets with RTP, UDP and IP headers. This encapsulation of call content presents serious technical challenges for a broadband access provider to identify and intercept VoIP calls.

First, broadband access service providers that simply transport IP traffic would encounter difficulty isolating VoIP-related RTP packets. Unlike SIP packets that are assigned to a specific port, RTP may use any port between 1,025 and 65,535 (although port 5,004 is often allocated as the default port when a port number is not otherwise allocated). Thus, it may be difficult for a broadband access provider to identify RTP traffic without either knowing the port number or "breaking open" and examining at least two layers of packet headers – something a broadband provider does not routinely do in the normal course of business.

Second, even if RTP packets could be isolated, broadband providers would have to decode and reassemble the voice data contained in the packet stream. RTP headers include a field identifying the encoding scheme used for the data contained in the packet payload, but simply knowing which protocol was used to encode a given piece of data is not enough. A broadband provider would also need to have the appropriate mechanisms for decoding the voice data, as well as access to the entire packet stream associated with a VoIP call.

As noted above, complications also exist in attempting to correlate the CII in the SIP packets with the associated VoIP call content in the RTP packets; and the broadband provider may not be able to do so without the help of the VoIP provider.

B. Complex Technical Solutions

In order to perform the above tasks, major new hardware and software would need to be implemented by broadband access service providers. Available options for doing so fall generally into the categories of packet filtering and traffic duplication, as discussed below. However, none of the available options provides a complete, cost-effective solution to the difficulties discussed above. It is particularly difficult for these solutions to address the problems presented by non-standard protocols, and the separation between VoIP content and signaling information.

1. Packet Filtering

One option for broadband access service providers is to implement routers with advanced packet filtering capabilities. But using such routers to identify source/destination IP addresses and the source/destination ports (to identify SIP traffic) would involve significant additional processing, with negative effects on network performance. For this reason, it is simply not practical at reasonable cost to use packet-filtering routers for high-speed, high-bandwidth IP traffic streams.

Packet filtering can also be implemented on a network-wide basis, using bandwidth management protocols such as DiffServ. DiffServ includes in IP packet headers information about the data being transmitted at higher network layers – *e.g.*, whether it is VoIP, streaming video, or something else.[15] Broadband providers can use DiffServ for bandwidth management (*i.e.*, giving privileged access to bandwidth to high-priority applications). However, bandwidth management protocols only work to the extent network edge hardware or software implements

[15] *See* K. Nichols, et al. RFC 2474 - Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers (1998), *available at* http://www.ietf.org/rfc/rfc2474.txt. *See also generally* http://www.ietf.org/html.charters/diffserv-charter.html.

the protocols. This may occur on private networks, or as part of a high-end VoIP service offered to private corporate customers. But it would be much more difficult for mass market broadband access providers to broadly implement bandwidth management services, since such service providers generally provide a generic IP-based Internet access service, and allow end-users to implement their own applications. It could have significant negative commercial impact if a provider were to force its customers to use only applications enabled for DiffServ or a similar protocol.[16]

2. Traffic Duplication

Alternatively, broadband access service providers could attempt to extract packets containing VoIP CII and call content from duplicated traffic. This approach could be implemented using a process known as port mirroring, or by adding a network tap device to the network.

Port mirroring (known as SPAN on Cisco equipment) is an advanced feature found on some switches whereby all traffic entering a particular port is duplicated and forwarded to a monitoring network connection. This allows a separate device to be connected to the switch and receive all of the duplicated traffic. Utilizing this approach, broadband access providers could then perform filtering operations on the duplicated traffic. However, this approach will not necessarily solve the performance degradation problems associated with filtering "live" traffic, as port mirroring can also have adverse affects on network performance. Moreover, port

[16] Furthermore, such providers may qualify for CALEA's private network exclusion; and even in the case of a provider of IP-enabled services to private corporate customers, it may be more efficient to apply intercepts within the enterprise rather than at some point in the provider's network.

mirroring often is unable to properly duplicate errors, such as undersized/oversized packets and framing errors. This could mean that not all traffic is actually sent to the monitoring device.

Another approach for intercepting CII is to install a tap – a separate network device for the sole purpose of duplicating traffic and sending it to a monitoring computer. A tap could be installed close to a switch that is used to process traffic sent to or from a targeted individual. If the tap possessed sufficient bandwidth capacity, this solution would be unlikely to degrade network performance, other than the potential downtime associated with installation. This is because instead of analyzing "live" traffic, filtering would be performed offline, on the duplicated packet stream. Furthermore, a tap can be combined with a packet filtering solution: commercial vendors currently offer off-the-shelf intelligent taps that can be used to monitor and identify a wide variety of network traffic, including VoIP.

Any approach based on processing of duplicated traffic risks imposing immense resource requirements on a broadband access provider. Since the duplicated traffic would not necessarily be limited to traffic destined for a particular IP address and/or port, the provider would essentially need to build a separate infrastructure to process all of the traffic passing through the traffic duplication device, with associated processing intelligence to analyze the packets for relevant CII that may be embedded within the packet using many different possible protocols. Furthermore, since such processing would be a resource-intensive activity that may be difficult to accomplish in real- or near real- time, the provider may also require significant amounts of data storage. One alternative to address these issues may be for duplicated traffic to be sent to trusted third parties ("TTPs") to filter and identify CII. However, this would not eliminate the technical burdens associated with "breaking open" packets, but simply shift them to third parties who may or may not be in a better position to spread the costs of such additional processing.

Moreover, interfaces with TTPs would need to be carefully developed to ensure network integrity and data security.

IV. TECHNICAL CHALLENGES THAT AFFECT VOIP SERVICE PROVIDERS

Although delivery of call content and CII of VoIP may be somewhat more straightforward for some VoIP service providers (*e.g.*, those that provide an end-to-end service and/or PSTN connectivity), this is not necessarily the case. Each VoIP service provider faces unique challenges, depending on the specific services it provides. For example, some VoIP service providers may do no more than provide SIP proxy services whereby users can look up other users and connect to one another. Such providers would never even see the ensuing VoIP traffic, thus precluding their ability to intercept call content which is carried on other providers' networks. Furthermore, even VoIP providers that have access to a user's VoIP communications would need to modify their systems, often significantly, in order to add intercept capabilities. This section details some of the difficulties faced by particular types of VoIP providers.

A. VoIP Over Broadband

Users often obtain broadband access and VoIP service from different providers (*e.g.*, Vonage over Comcast broadband). In this situation, the broadband provider transports the SIP packets to the VoIP provider for call management. For IP-PSTN calls that the VoIP provider completes, the broadband provider would transport the RTP packets to the VoIP provider for interconnection with the PSTN. In such circumstances, the VoIP provider may be the more appropriate (and more efficient) entity to conduct the intercept of VoIP CII and call content. For calls that do not touch the PSTN, however, the VoIP provider may be involved in call setup but may or may not see the call content depending on whether the RTP packets are transported by its network or by the networks of others. Furthermore, while the VoIP provider would typically have access to signaling information used for call setup, additional signaling functions performed during the call may be performed entirely by customer software or equipment.

Furthermore, in order to conserve limited IP address resources, many broadband providers use dynamic IP addresses, which are assigned to whichever customers are active at a particular time. As a result, a VoIP provider may only be able to identify the broadband provider and not the actual customer of that broadband provider.

B. SIP Proxy Providers

A VoIP provider may do no more than provide SIP proxy services that allow users to establish a call. Call content would then be encapsulated as RTP packets and transmitted as IP traffic over other providers' network. In this situation, it is doubtful whether the VoIP provider would have any access to the call content.

C. VoIP Wholesale and Resale

The user's VoIP provider may perform some signaling functions before handing off the traffic to a wholesale VoIP provider for transport (*e.g.*, a VoIP-optimized backbone service provider). Or the user's VoIP provider may be a pure reseller of another provider's wholesale VoIP service. In these circumstances, and in addition to the issues identified above, no one service provider would be able to provide all call content and/or CII without cooperation from other providers.

V. SUMMARY AND CONCLUSION

In sum, the technical challenges associated with the extraction of CII and call content are significant, especially when multiple providers in the service supply chain handle different aspects of the VoIP service. In a multi-provider environment, no single provider is likely to be able to easily access all relevant information that may be of interest to law enforcement. Service providers would have to significantly modify their networks, and implement complicated technical solutions, to "break open" packets to extract information from parts of the packet they would not use during the normal course of business. The better solution may not be to require any one provider to implement costly and complicated methods for extracting the relevant information, but to have law enforcement request the information from the provider that has easiest access to it.